

# Aligning Pharmacologic Classes Between MeSH and ATC

Rainer Winnenburger, Laritza Taft, Fiona Callaghan, and Olivier Bodenreider<sup>\*</sup>  
Lister Hill National Center for Biomedical Communications, National Library of Medicine, Bethesda, USA

## ABSTRACT

**Objective:** To align pharmacologic classes in ATC and MeSH with lexical and instance-based techniques.

**Methods:** Lexical alignment: we map the names of ATC classes to MeSH through the UMLS, leveraging normalization and additional synonymy. Instance-based alignment: we associate ATC and MeSH classes through the drugs they share, using the Jaccard coefficient to measure class-class similarity. We use a metric to distinguish between equivalence and inclusion mappings.

**Results:** We found 226 lexical mappings, as well as 360 instance-based mappings, with a limited overlap (62). From the 360 instance-based methods we classify 113 as equivalence mappings and 247 as inclusion mappings. A limited failure analysis is presented.

**Conclusion:** Our instance-based approach to aligning pharmacologic classes has the prospect of effectively supporting the creation of a mapping of pharmacologic classes between ATC and MeSH. This exploratory investigation needs to be evaluated in order to adapt the thresholds for similarity.

## 1 INTRODUCTION

The National Library of Medicine (NLM) and the Food and Drug Administration (FDA) Center for Drug Evaluation and Research (CDER) are collaborating on a research project to extract adverse drug reactions from the biomedical literature. More specifically, this investigation leverages the indexing of MEDLINE citations to extract associations between drug entities and clinical manifestations in the context of adverse events.

Since the biomedical literature is indexed with the Medical Subject Headings (MeSH) vocabulary, this data mining effort essentially consists in identifying salient co-occurrences of MeSH terms for drug entities and for manifestations. However, adverse drug reactions are usually analyzed not in reference to MeSH, but in reference to other standard vocabularies, namely the Anatomical Therapeutic Chemical (ATC) drug classification system for drug entities, and the Medical Dictionary for Regulatory Activities (MedDRA) for clinical manifestations.

In this project, drug entities have to be mapped from MeSH to ATC, and manifestations from MeSH to MedDRA. Drug entities include not only individual drugs (e.g., *atorvastatin*), but also drug classes (e.g., *statins*). In previous work, we have mapped individual drugs between RxNorm (which includes MeSH drugs) and ATC (Bodenreider and Taft, 2013; Winnenburger and Bodenreider, 2012). In contrast, no mapping is available between pharmacologic classes in

MeSH and in ATC. Moreover, unlike individual drugs, whose names are relatively standardized across vocabularies, pharmacologic classes exhibit greater variability, not only in their names, but also in granularity. For example, the drug *lisinopril* is classified as *Angiotensin-Converting Enzyme Inhibitors* in MeSH, but as *ACE inhibitors, plain* in ATC.

The objective of this study is to investigate ontology matching techniques for aligning pharmacologic classes between MeSH and ATC. Such methods are expected to facilitate the curation of a mapping by experts by identifying candidate pairs of equivalent classes in MeSH and ATC. To our knowledge, this work represents the first effort to map pharmacologic classes between MeSH and ATC.

## 2 BACKGROUND

The general framework of this study is that of ontology alignment (or ontology matching). Various techniques have been proposed for aligning concepts across ontologies, including lexical techniques (based on the similarity of concept names), structural techniques (based on the similarity of hierarchical relations), semantic techniques (based on semantic similarity between concepts), and instance-based techniques (based on the similarity of the set of instances of two concepts). An overview of ontology alignment is provided in (Euzenat and Shvaiko, 2007).

The main contribution of this paper is not to propose a novel technique, but rather to apply existing techniques to a novel objective, namely aligning pharmacologic classes between MeSH and ATC. To this end, we use lexical and instance-based techniques, because the names of pharmacologic classes and the list of drugs that are members of these classes are the main two features available in these resources.

### 2.1 Lexical techniques

Lexical techniques for ontology matching compare concept names across ontologies. When synonyms are available, they can be used to identify additional matches. Matching techniques beyond exact match utilize edit distance or normalization to account for minor differences between concept names.

As part of the Unified Medical Language System (UMLS), linguistically-motivated normalization techniques have been developed specifically for biomedical terms (McCray, et al.,

<sup>\*</sup> To whom correspondence should be addressed: obodenreider@mail.nih.gov

1994). UMLS normalization abstracts away from inessential differences, such as inflection, case and hyphen variation, as well as word order variation. The UMLS normalization techniques form the basis for integrating terms into the UMLS Metathesaurus, but can be applied to terms that are not in the UMLS. For example, the ATC class *Thiouracils* (H03BA) and the MeSH class *Thiouracil* (D013889) match after normalization (ignoring singular/plural differences).

Lexical techniques typically compare the names of concepts across two ontologies as provided by these ontologies. However, additional synonyms can be used, for example, synonyms from the UMLS Metathesaurus. In other words, we leverage cosynonymy similarity for matching pharmacologic classes. In this case, although the ATC class *Anticholinesterases* (N06DA) and the MeSH class *Cholinesterase Inhibitors* (D002800) do not match lexically, both names are cosynonyms, because they are found among the synonyms of the UMLS Metathesaurus concept C0008425.

## 2.2 Instance-based techniques

Also called extensional techniques, instance-based techniques compare classes based on the sets of individuals (i.e., instances) of each class. Many biomedical ontologies consist of class hierarchies, but do not contain information about instances. Here, however, individual drugs (e.g., *atorvastatin*) are the members – not subclasses – of pharmacologic classes (e.g., *statins*). In other words, pharmacologic classes have individual drugs as instances, not subclasses.

Several methods have been proposed to implement instance-based matching. (Isaac, et al., 2007) decompose these methods into three basic elements:

- A measure is used for evaluating the association between two classes based on the proportion of shared instances. Typical measures include information-based measures (e.g., Jaccard similarity coefficient) and statistical measures (e.g., log likelihood ratio).
- A threshold is applied to the measures and pairs of classes for which the measure is above the threshold are deemed closely associated and mapping candidates.
- Hierarchical relations in the two ontologies to be aligned can also be leveraged by deriving instance-class relations between instances of a given class and the ancestors of this class. In other words, in addition to asserted classes (i.e., the classes of which individual drugs are direct members), we also consider inferred classes (i.e., the classes of which asserted classes are subclasses). For example, the class asserted in MeSH for the drug *atorvastatin* is *Hydroxymethylglutaryl-CoA Reductase Inhibitors* (i.e., *statins*), whose parent concepts include *Anticholesteremic Agents*. Therefore, the class *Anticholesteremic Agents* is an inferred pharmacologic class for *atorvastatin*.

## 2.3 Related work

Lexical techniques are a component of most ontology alignment systems (Euzenat and Shvaiko, 2007). While instance-based techniques are also available in many systems, the applicability of this technique is limited, because there is often no available information about instances as part of the ontologies to be aligned. For example, most biomedical terminologies and ontologies are simple class hierarchies. The instances of these classes are present in electronic medical record systems and clinical data warehouses, but typically not distributed along with the ontologies. One exception in the biomedical domain is the Gene Ontology (Ashburner, et al., 2000), for which the gene products annotated to Gene Ontology terms can be considered instances of the corresponding classes. (Kirsten, et al., 2007) have aligned GO terms across the three hierarchies of GO through the genes products to which they are co-annotated.

To our knowledge, our work is the first attempt to align pharmacologic classes with instance-based techniques (i.e., beyond name matching), and the first application of alignment between pharmacologic classes in ATC and MeSH. Moreover, while most ontology alignment systems mainly consider matches between equivalent classes, we are also interested in identifying those cases where one class is included in another class.

## 3 MATERIALS

### 3.1 Anatomical Therapeutic Chemical Drug Classification System (ATC)

The ATC is a clinical drug classification system developed and maintained by the World Health Organization (WHO) as a tool for drug utilization research to improve quality of drug use (ATC, 2013). The system is organized as a hierarchy that classifies clinical drug entities at five different levels: 1st level anatomical (e.g., *A: Alimentary tract and metabolism*), 2nd level therapeutic (e.g., *A10: Drugs used in diabetes*), 3rd level pharmacological (e.g., *A10B: Blood glucose lowering drugs, excluding insulins*), 4th level chemical (e.g., *A10BA: Biguanides*), and 5th level chemical substance or ingredient (e.g., *A10BA02: metformin*). The 2013 version of ATC integrates 4,516 5<sup>th</sup>-level drugs and 1,269 drug groups (levels 1-4).

### 3.2 MeSH

The Medical Subject Headings (MeSH) is a controlled vocabulary produced and maintained by the NLM (NLM, 2013). It is used for indexing, cataloging, and searching the biomedical literature in the MEDLINE/PubMed database, and other documents. The MeSH thesaurus includes 26,853 descriptors (or “main headings”) organized in 16 hierarchies (e.g., *Chemical and Drugs*). Additionally, MeSH provides about 210,000 supplementary concept records (SCRs),

many of which represent chemicals and drugs. Each SCR is linked to at least one descriptor. While most chemical descriptors provide a structural perspective on drugs, some descriptors play a special role as they can be used to denote pharmacological actions in drug descriptors and SCRs. MeSH 2013 is used in this study.

### 3.3 RxNorm

RxNorm is a standardized nomenclature for medications produced and maintained by the U.S. National Library of Medicine (NLM) (NLM, 2013). RxNorm concepts are linked by NLM to multiple drug identifiers for commercially available drug databases and standard terminologies, including MeSH. RxNorm serves as a reference terminology for drugs in the US. The March 2013 version of RxNorm used in this study integrates about 10,500 base and salt ingredients. NLM also provides an application programming interface (API) for accessing RxNorm data programmatically (NLM, 2013).

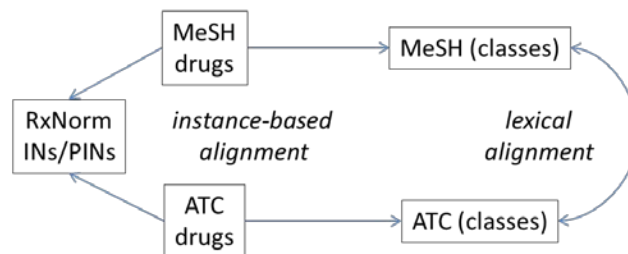
### 3.4 Unified Medical Language System (UMLS)

The UMLS is a terminology integration system created and maintained by the National Library of Medicine (NLM) (NLM, 2013). The UMLS Metathesaurus integrates over 150 terminologies, including MeSH, but not ATC. Synonymous terms across terminologies are grouped into concepts and assigned the same concept unique identifier. The Metathesaurus provides a comprehensive set of synonyms for biomedical concepts and is often used for integrating terminologies beyond its own. NLM provides an application programming interface (API) for accessing UMLS data programmatically. Version 2012AB of the UMLS is used in this study.

## 4 METHODS

Our approach to aligning pharmacologic classes between MeSH and ATC based on their instances is depicted in Figure 1 and can be summarized as follows. We started our investigation by establishing a lexical alignment of MeSH and ATC classes based on the class names and their synonyms (Figure 1, right). We then constructed an instance-based alignment of MeSH and ATC classes considering the individual drugs shared by the classes (Figure 1, left). We mapped individual drugs from MeSH and ATC via their ingredients (IN) or precise ingredients (PIN) in RxNorm. We used a similarity measure and thresholds to identify class mappings and compared them with the mappings retrieved by the lexical approach.

In our alignment work, we excluded 124 of the 1,255 ATC groups ( $2^{\text{nd}}$  –  $4^{\text{th}}$  level) corresponding to drug combinations, because combination drugs are often underspecified in ATC. We also excluded the 14 ATC groups of level 1 (anatomical classification), because they are too broad classes.



**Figure 1.** Alignment of ATC and MeSH classes, alignment via their instances (left) in comparison to direct lexical mapping of the class names (right).

Similarly, in MeSH, we excluded the top-level descriptors of the Chemicals and Drugs hierarchy (i.e., D01 - D27), as well as the top-level of the pharmacological action descriptors (*Pharmacologic Actions*, *Molecular Mechanisms of Pharmacological Action*, *Physiological Effects of Drugs*, and *Therapeutic Uses*).

#### 4.1 Lexical alignment

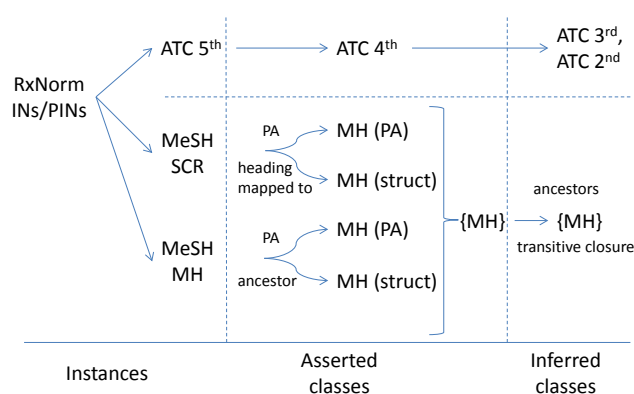
We mapped all 1,117 eligible ATC classes ( $2^{\text{nd}}$  –  $4^{\text{th}}$  level) to MeSH descriptors in the Chemicals and Drugs [D] tree using the UMLS Terminology Services (UTS). More precisely, we used the *ExactString* and *NormalizedString* search function of the UTS API 2.0 to establish mappings from the names of the ATC classes to UMLS concepts. We used normalization only when the exact technique did not result in a mapping. We then mapped the UMLS concepts to MeSH descriptor IDs.

#### 4.2 Instance-based alignment

**Mapping ATC drugs to RxNorm ingredients.** In previous work we have mapped ATC single-ingredient drugs to Ingredients (IN) and Precise Ingredients (PIN) in RxNorm using a lexical approach with additional normalization steps (Winnenburg and Bodenreider, 2012). We used these mappings to establish the alignment of ATC and RxNorm drugs in this study.

**Mapping MeSH drugs to RxNorm ingredients.** Since MeSH drugs are integrated in RxNorm, mappings to equivalent drug concepts from MeSH can be obtained via the *getProprietaryInformation* function from the RxNorm API. We systematically exploited this information for all Ingredients (IN) and Precise Ingredients (PIN) in RxNorm and created a mapping table between RxNorm CUIs and MeSH Main Headings (MH) and Supplementary Concept Records (SCR).

**Inferring class membership in ATC.** We considered the hierarchical relations from  $5^{\text{th}}$  level drugs to their  $4^{\text{th}}$  level chemical groups as asserted drug class membership (Figure 2, top). We inferred membership between  $5^{\text{th}}$  level drugs and groups of level 3 and 2 through transitive closure.



**Figure 2.** Mapping ATC and MeSH drugs via RxNorm ingredients and inferring class memberships in ATC and MeSH.

**Table 1.** Asserted and inferred MeSH classes for the drug *temafloxacin* (C054745) with type of relationship to the drug and tree numbers in MeSH.

Type	Asserted Classes	Inferred Classes
PA	<i>Anti-Bacterial Agents</i> (D000900) [D27.505.954.122.085]	<i>Anti-Infective Agents</i> (D000890) [D27.505.954.122]
MH	<i>Fluoroquinolones</i> (D024841) D03.438.810.835.322	<i>Quinolones</i> (D015363) [D03.438.810.835]  <i>Quinolines</i> (D011804) [D03.438.810]  <i>Heterocyclic Compounds, 2-Ring</i> (D006574) [D03.438]

For example, *temafloxacin* (J01MA05) is a member of the chemical group *Fluoroquinolones* (J01MA - asserted), the pharmacological group *QUINOLONE ANTIBACTERIALS* (J01M - inferred), and the therapeutic group *ANTIBACTERIALS FOR SYSTEMIC USE* (J01 - inferred).

**Inferring class membership in MeSH.** We associated each RxNorm drug (IN or PIN) with its corresponding MeSH SCR or MH. In turn, we associated these drugs with their asserted classes as shown in Figure 2. For an SCR, we considered its pharmacological actions, as well as the MeSH headings mapped to. For an MH, we considered its pharmacological actions, as well as its direct ancestors. These constitute the asserted classes. We inferred membership between the drugs and higher-level descriptors in the MeSH hierarchy. For example, as shown in Table 1, the SCR *temafloxacin* has *Anti-Bacterial Agents* as pharmacological action and *Fluoroquinolones* as main heading mapped to. From these asserted classes, we infer membership to *Anti-Infective Agents* (from *Anti-Bacterial Agents*) and to *Quinolones*, *Quinolines*, and *Heterocyclic Compounds, 2-Ring* (from *Fluoroquinolones*).

*olones*, *Quinolines*, and *Heterocyclic Compounds, 2-Ring* (from *Fluoroquinolones*).

**Measure for aligning ATC and MeSH classes.** Based on the asserted and inferred class membership of drugs in ATC and MeSH we conducted a pairwise comparison of all ATC against all MeSH classes. For each pair of ATC class (A) and MeSH class (M), we computed the Jaccard coefficient. In order to reduce the similarity of pairs of classes with a small number of shared members, we used a modified version of the Jaccard coefficient, JMod, as suggested in (Isaac, et al., 2007),

$$JC(A, M) = \frac{|A \cap M|}{|A \cup M|}$$

$$JMod(A, M) = \frac{\sqrt{|A \cap M| \times (|A \cap M| - 0.8)}}{|A \cup M|}$$

where  $A \cap M$  represents the number of drugs common to A and M, and  $A \cup M$  the total number of unique drugs in both classes.

The Jaccard coefficient measures the similarity between the two classes, but does not reflect whether one class is included in the other. Because of the difference in granularity between classes in ATC and MeSH, we introduce a simple metric for detecting whether the drugs that are not shared by both classes are primarily in one of the two classes. This “one-sidedness” coefficient is calculated as follows:

$$0, \quad \text{for } a = 0 \text{ and } m = 0$$

$$|a-m| / a+m, \quad \text{otherwise.}$$

where a and m are the number of drugs specific to the ATC class and the MeSH class, respectively.

**Thresholds.** In order to select the best equivalent or inclusion mappings between ATC and MeSH, we characterize each pair of ATC and MeSH classes with respect to Jaccard similarity and one-sidedness. Low one-sidedness indicates equivalence and high one-sidedness indicates inclusion. High Jaccard similarity indicates strong overlap between the two classes. Based on preliminary analysis, we selected of a threshold of 0.5 for the one-sidedness metric. Similarly, we selected of a threshold of 0.5 and 0.25 for Jaccard similarity for equivalence (low one-sidedness) and inclusion (high one-sidedness), respectively. The lower threshold for Jaccard similarity for inclusion was determined empirically. As shown in Table 2, each pair of ATC and MeSH classes is characterized as an equivalence mapping (EQ+), an inclusion mapping (IN+), or not a mapping (EQ- and IN-).

## 5 RESULTS

### 5.1 Lexical alignment

For the 1,117 eligible ATC groups of level 2-4, we were able to retrieve 226 mappings to descriptors from the Chemicals and Drugs [D] tree in MeSH. We have 18 mappings for therapeutic classes (2<sup>nd</sup> level), 42 for pharmacological classes (3<sup>rd</sup> level), and 166 for chemical classes (4<sup>th</sup> level). We ignored mappings for the broad anatomical classes (1<sup>st</sup> level). Of the 226 mappings, 99 are to pharmacological actions (functional perspective) in MeSH, whereas 127 are to other descriptors at various levels of the MeSH hierarchy (structural perspective).

### 5.2 Instance-based alignment

Of the 1,117 eligible ATC groups, 874 (78%) could be associated with at least one descriptor or pharmacological action in MeSH. We identified a total of 930 associations for the 874 ATC groups. As shown in Table 2, based on the one-sidedness metric, we characterized 322 associations as equivalence and 608 as inclusion. Of the 322 equivalence associations, 113 (35%) exhibit high Jaccard similarity and are selected as equivalence mappings (EQ+). Of the 608 inclusion associations, 247 (41%) exhibit high Jaccard similarity and are selected as inclusion mappings (IN+). The other associations (EQ- and IN-) are not deemed strong enough to denote mapping. In summary, we were able to characterize as a mapping (EQ+ and IN+) 360 (39%) of the associations between ATC and MeSH classes. It should be mentioned that we were not able to obtain mappings to MeSH classes for 243 ATC classes, because they only contain drug instances that could not be mapped to drugs in MeSH.

**Table 2.** Characterization of the associations between ATC and MeSH classes based on Jaccard similarity and score for one-sidedness.

		One-sidedness		
		≥ .5	< .5	Total
Jaccard	≥ .5	IN+ (247)	EQ+ (113)	360
	[.25-.5[		EQ- (209)	570
	< .25	IN- (361)		
	Total	608	322	930

### 5.3 Comparison between lexical and instance-based alignment

As illustrated in Table 3, from the 226 lexical mappings between ATC and MeSH classes, we could confirm 62 with our instance-based approach (30 as equivalence mappings, 32 as inclusion mappings). For 16 of the lexical mappings

we found an association with low Jaccard similarity (IN- / EQ -), and for 148 we did not find any association through the instance-based alignment. Finally, the instance-based approach produced 298 drug class additional mappings that were not detected by the lexical approach.

**Table 3.** Comparison between lexical and instance-based alignment.

		Instance-based			Total
		Yes	No	No assoc.	
Lexical	Yes	62	16	148	226
	No	298	554		852
Total		360	570	148	1078

## 6 DISCUSSION

### 6.1 Examples and failure analysis

**True positive for equivalent instance-based mappings.** We identify an equivalence mapping between the 4<sup>th</sup>-level ATC group *Fluoroquinolones* (J01MA) and the MeSH descriptor *Fluoroquinolones* (D024841). The two classes share 14 drugs. The ATC group has one extra drug (*moxifloxacin*), and the MeSH descriptor has 2 (*flumequine* and *besifloxacin*). Jaccard similarity is high (0.82) and the one-sidedness score is low (0.33), because the 3 drugs that are not in common are not all on the same side. This mapping is also identified by the lexical technique (exact match).

**True positive for inclusion instance-based mappings.** We identify an inclusion mapping between the 4<sup>th</sup>-level ATC group *Fluoroquinolones* (S01AE) and the MeSH descriptor *Fluoroquinolones* (D024841). Although the two classes are seemingly identical, our mapping is identified as an inclusion, with 7 drugs in common, 1 drug specific to the ATC class and 9 drugs specific to the MeSH class. In fact, the ATC class is not the same general class for anti-infective agents as in the example above (J01MA), but rather the specific class of fluoroquinolones for ophthalmic use (S01AE). The fluoroquinolones used for eye disorders are a subset of all fluoroquinolones and the ATC class S01AE is appropriately characterized as being included in the MeSH class for fluoroquinolones. This example also illustrates a false positive for the lexical mapping, since it is generally assumed that lexical mappings are equivalence mappings.

**False negative for equivalent instance-based mappings.** Many ATC and MeSH classes share only one of very few drugs, making it difficult to assess equivalence or inclusion. For example, the 4<sup>th</sup>-level ATC group *Silver compounds* (D08AL) and the MeSH descriptor *Silver Compounds* (D018030) share only one drug (silver). The modified version of the Jaccard coefficient has a score of 0.45 in this case, which is below our threshold of 0.5 for equivalence.



During this failure analysis, we discovered that some MeSH drugs did not have a pharmacological action assigned to them as we expected. For example, while *pyrantel* is listed as *Antinematodal Agents*, *oxantel* is not. We are investigating whether the pharmacological action for this SCR should be inferred from the descriptor to which it is mapped (*Pyrantel* in this case). Because of these missing pharmacologic actions, the 3<sup>rd</sup>-level ATC group *ANTINEMATODAL AGENTS* (P02C) fails to be mapped to the MeSH pharmacological action *Antinematodal Agents* (D000969), the Jaccard similarity being just below the threshold (0.49).

**Discrepancy between lexical and instance-based alignment (missed lexical mapping).** Despite the use of UMLS synonymy and normalization, the lexical alignment fails to identify a mapping between the 3<sup>rd</sup>-level ATC group *POTASSIUM-SPARING AGENTS* (C03D) and the MeSH pharmacological action *Diuretics, Potassium Sparing* (D062865). In contrast, the instance-based alignment identifies an equivalence mapping with very high Jaccard similarity (0.99).

**Discrepancy between lexical and instance-based alignment (missed instance-based mapping).**

We have identified several causes for discrepancies between lexical and instance-based alignments. As mentioned earlier, some ATC classes only contain drugs that cannot be mapped to MeSH through RxNorm, which we used to bridge between the two. Sometimes, the best instance-based mapping is to another class than the class found by the lexical technique. Finally, some drugs entities and biologicals (e.g., vaccines) are less well standardized than common drugs. For this reason, the instance-based alignment is unable to map these classes, when simple lexical techniques can.

## 6.2 Limitations and future work

This exploratory investigation has several limitations, which we plan to address in future work.

**Evaluation.** We did not perform any formal evaluation of the alignment. Creating a gold standard requires domain experts and is labor intensive. This preliminary work will help us establish the guidelines for creating such a gold standard.

**Perspective.** Our perspective in this investigation is ATC-centric, because we consider the best MeSH mapping for each ATC class, but not the best ATC mapping for each MeSH class. One future goal is to explore both directions using the same methodology.

**Bias towards equivalence mappings.** Because we restrict our exploration to the MeSH class with the best Jaccard similarity for each ATC class (which we subsequently categorize as equivalence or inclusion), and because of the differential threshold for Jaccard similarity between equivalence

(0.5) and inclusion mappings (0.25), we potentially fail to consider a good inclusion mapping (e.g., with a similarity score of 0.39 [ $> 0.25$ ]), when the best MeSH class is a bad equivalent mapping (e.g., with a similarity score of 0.41 [ $< 0.5$ ]).

## 6.3 Significance

To our knowledge, our work is the first attempt to align pharmacologic classes with instance-based techniques, as well as the first application of alignment between pharmacologic classes in ATC and MeSH. Our instance-based approach to aligning pharmacologic classes has yielded 360 mappings, and has the prospect of effectively supporting the creation of a mapping of pharmacologic classes between ATC and MeSH. This exploratory investigation needs to be evaluated in order to adapt the thresholds for similarity.

## ACKNOWLEDGEMENTS

This work was supported by the Intramural Research Program of the NIH, National Library of Medicine.

## REFERENCES

- Ashburner, M., et al. (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium, *Nat Genet*, **25**, 25-29.
- Anatomical Therapeutic Chemical (ATC) classification: <http://www.whocc.no/atc/>
- Bodenreider, O. and Taft, L.M. (2013) A mapping of RxNorm to the ATC/DDD Index helps analyze US prescription lists, *AMIA Annu Symp Proc*, (submitted).
- Euzenat, J. and Shvaiko, P. (2007) *Ontology matching*. Springer, New York.
- Isaac, A., et al. (2007) An empirical study of instance-based ontology matching. In Aberer, K., et al. (eds), *Proceedings of the 6th international The semantic web and 2nd Asian conference on Asian semantic web conference (ISWC'07/ASWC'07)*. Springer-Verlag, pp. 253-266.
- Kirsten, T., Thor, A. and Rahm, E. (2007) Instance-based matching of large life science ontologies In Cohen-Boulakia, S. and Tannen, V. (eds), *Data Integration in the Life Sciences: 4th International Workshop, DILS 2007, Philadelphia, PA, USA*. Springer, pp. 172-187.
- McCray, A.T., Srinivasan, S. and Browne, A.C. (1994) Lexical methods for managing variation in biomedical terminologies, *Proc Annu Symp Comput Appl Med Care*, 235-239.
- Medical Subject Headings (MeSH): <http://www.nlm.nih.gov/mesh/>
- RxNorm: <http://www.nlm.nih.gov/research/umls/rxnorm/>
- RxNorm API: <http://rxnavdev.nlm.nih.gov/RxNormAPI.html>
- Unified Medical Language System (UMLS): <https://uts.nlm.nih.gov/>
- Winnenburg, R. and Bodenreider, O. (2012) Mapping drug entities between the European and American standards, ATC and RxNorm, *Poster Proceedings of the Eighth International Conference on Data Integration in the Life Sciences (DILS 2012)*, 22.